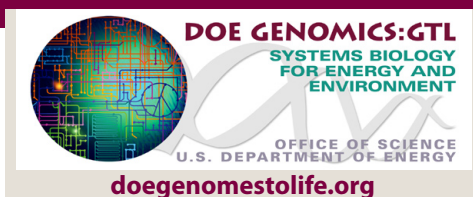


Genomics:GTL *Systems Biology for Energy and Environment* Program Overview

The growing complexity of U.S. Department of Energy (DOE) mission challenges in energy security, environmental restoration, and climate protection requires the development of breakthrough technologies. In the coming decades, biology-based technologies using living organisms or their components such as genes and proteins can play a growing role in providing innovative solutions and in informing policy and decision making.

The Genomics:GTL (GTL, formerly Genomes to Life) research program focuses on developing technologies to understand and use the diverse capabilities of microbes, which make up the foundation of the biosphere and sustain all life on earth (see *The Microbial World*, pp. 2–3). DOE has sponsored the genome sequencing of some 200 microbes relevant to generating clean energy, cleaning up toxic waste from nuclear weapons development, and cycling carbon from the atmosphere. Before we can harness and use microbial processes, however, they must be understood in far greater detail and in the realistic context of living, dynamic systems—whether as individual cells or communities of interacting cells—rather than such components as genes or proteins.

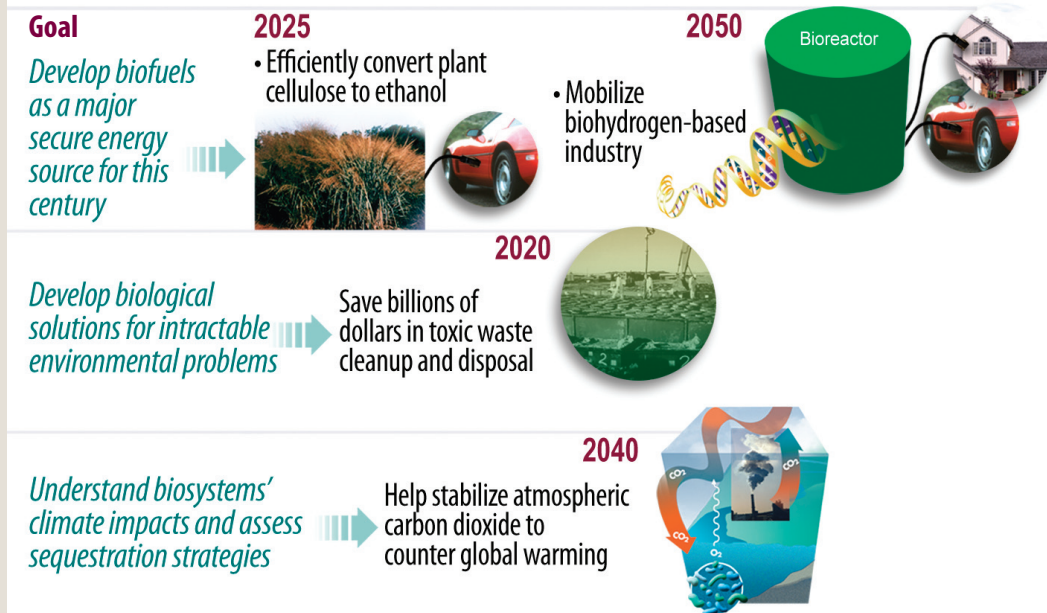


Achieving a Predictive Understanding Through Systems Biology

The ultimate GTL goal is to generate increasingly accurate mathematical models of life processes that allow the use of genome (DNA) sequence to predict microbial cell and community responses to varying environmental conditions. This level of knowledge will support the creation and confident use of new or modified systems tailored for specialized applications.

A comprehensive approach to understanding biological systems must encompass genes and proteins, multimolecular assemblies (“molecular machines”), pathways and interacting networks, whole cells, communities of cells, and environments (see sidebar, *The Basics*, p. 4). Surmounting the technical challenges presented by these multiscale explorations into whole living systems is a daunting prospect that will require dramatic improvements in research performance, throughput, quality, and cost. New capabilities also will be needed in computation, modeling, and simulation, all of which are integral parts of systems biology research. The GTL research strategy (p. 4) is designed to meet these challenges.

Grand Challenges for Biology, Payoffs for the Nation



Catalyzing Research and Industry

GTL will make its resources, facilities, and knowledgebase available to all scientists and industry, enabling cutting-edge investigations on the systems biology level and fostering participation by the greater community in solving DOE mission problems. These enabling capabilities also will facilitate rapid translation of science into new technologies and catalyze the industrial biotechnology sector of the economy.

THE MICROBIAL WORLD:

A Vast and Genetically Rich Resource

Microbes and their communities make up the foundation of the biosphere and sustain all life on earth. These single-celled organisms are masters at living in almost every environment and harvesting energy in almost any form, from solar radiation to photosynthesis-generated organic chemicals to minerals in the deep subsurface.

Microbes have evolved over 3.5 billion years, transforming the atmosphere with oxygen (a by-product of photosynthesis) more than a billion years ago to create the environment for life as we know it. Some microbes can thrive in either aerobic (with oxygen) or anaerobic (without oxygen) conditions. Microbes also capture nitrogen from the atmosphere, make it available to plants (and other life forms), and carry out processes responsible for soil fertility. Most do not cause disease. The unique microbial biochemistries amassed over eons in every niche on the planet now offer a deep and virtually limitless resource of capabilities that can be applied to national needs, including DOE energy and environmental missions.

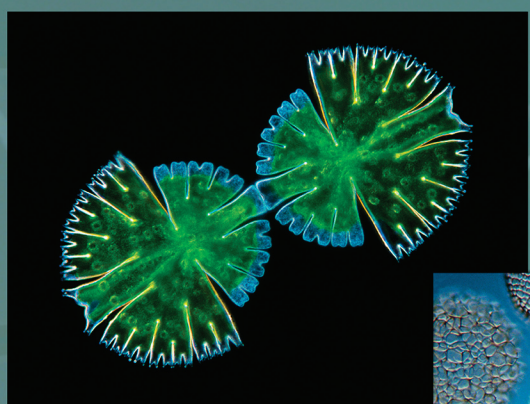
Although immense, the microbial world remains largely unexplored, a frontier of truly astronomical dimensions: The estimated nonillion or 10^{30} individual bacteria on earth are 10^9 times more than the number of stars in the universe. The vast majority, however, cannot be studied using standard techniques. While 2000 to 3000 species are estimated to be present in a single gram of soil, we can cultivate for study only some 0.1 to 1% of the species in that or any other environment. About 5700 species have been described thus far.¹⁻³

Investigators now are beginning to apply the tools of genomics to studying this enormous untapped natural treasure. Because microbes have modest-sized genomes (averaging 4 to 5 million bases compared with 3 billion bases in the human and other mammalian genomes), they represent a tractable life form we can use to explore and understand life processes at a whole-system level. Already, limited environmental sampling of microbes and their communities has led to the

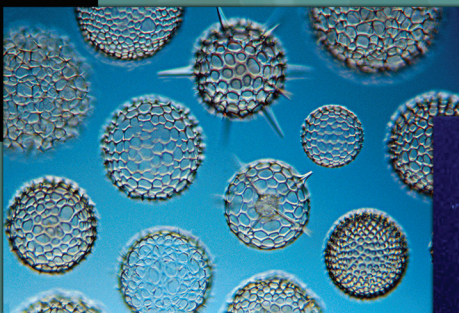
discovery of millions of previously unknown genes and proteins, thousands of species, and innumerable variations in critical functionalities. As scientists begin to scratch the surface of the microbial world, they are finding analysis an enormous challenge.

Recent discoveries from projects funded by DOE's Biological and Environmental Research program highlight the ubiquitous presence and critical importance of microbes in all ecosystems. For example:

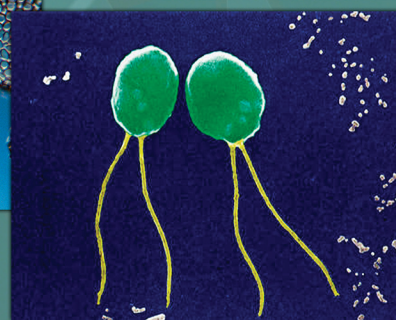
■ The cyanobacteria *Prochlorococcus* and *Synechococcus*, along with other ocean phytoplankton, account for about half of global photosynthesis.⁴



***Micrasterias rotata*, a Desmid Undergoing Cell Division or Cytokinesis.** Desmids are attractive unicellular freshwater green algae that have a distinct basic symmetry consisting of two semicells with the nucleus situated in the narrow center. When they divide, two new semicells are formed.
[© Wim van Egmond / Visuals Unlimited]



***Spumellarian radiolarian*, Skeletons from the Ocean Bottom.** Radiolarians are unicellular protists with strikingly beautiful siliceous skeletons showing radial symmetry.
[© Wim van Egmond / Visuals Unlimited]



***Chlamydomonas*, Green Algae with Two Flagella for Movement.** These microbes can generate hydrogen from light, water, and basic nutrients.
[Elias Greenbaum, Oak Ridge National Laboratory]

- Diatoms, ancient and intricately shaped ocean microbes, store an amount of carbon comparable to that in all the earth's rainforests combined. Over geological time, diatoms may have influenced the earth's climate.⁵
- More than a million previously undiscovered genes, possibly representing new biochemical functions, were the surprising find in sequencing DNA fragments from the Sargasso Sea—a region heretofore thought to sustain little life.⁶ This discovery also was named one of *Science* magazine's "Breakthroughs of the Year."⁷

A CHALLENGING FRONTIER

- Microbes thrive deep within the earth's subsurface and at extremes previously thought to extinguish life.⁸

Growing recognition of microbial capabilities and potential applications has made a compelling case for further investigations by DOE and other agencies and institutions.

Before we can harness their capabilities, microbes must be understood in far greater detail and in the realistic context of whole living systems—whether as individuals or communities of interacting microbes—rather than as isolated components such as single genes and proteins. Microbes already can be manipulated at the molecular, cellular, and system levels, but understanding and taking advantage of their complexities and surmounting the technical challenges of whole-systems biology is a daunting prospect.

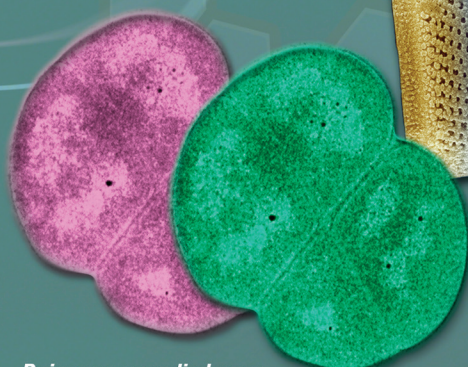
Understanding MICROBES and Their Communities

Most microbes live in highly organized and interactive communities that are versatile, complex, and difficult to analyze from many perspectives. Some of these challenges are outlined below.

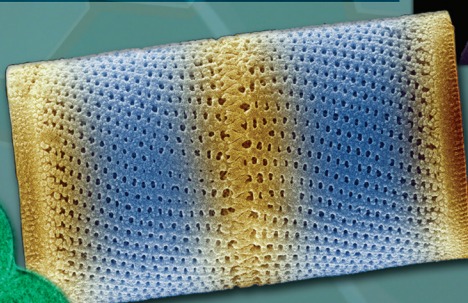
- Microbes are exceedingly small—only 1/8000th the volume of a human cell and spanning about 1/100th the diameter of a human hair. Investigating processes within this size range is challenging.
- The microbial world encompasses millions of genes from thousands of species, with hundreds of thousands of proteins and multimolecular machines operating in a web of hundreds of interacting processes in response to numerous physical and chemical environmental variables. Gene control is complex, with groups or “cassettes” of genes (operons) directing coordinated transcription and translation of genes into interacting proteins.
- Microbes adapt rapidly in response to environmental change, an ability that underlies their survival for billions of years. For example, various species of “extremophile” microbes have adapted to great extremes of pressure, temperature, pH, salinity, and radiation. Their high surface-to-volume ratio enhances interactions and supports adaptation. Unlike animal cells, they have no protective nucleus for their DNA, which leaves it more vulnerable to alteration. Genes move easily among species. Moreover, microbial communities are awash in genetic material from viruses that confer additional genetic properties and expand their range of adaptability.



Rod-Shaped (Bacilli) and Spherical (Cocci) Bacteria Found in Compost. Decomposition of organic matter is an extremely important process in nature and a part of the global carbon and nutrient cycle.
[© Simko / Visuals Unlimited]



***Deinococcus radiodurans*, the Most Radiation-Resistant Microbe Known.**
[Michael Daly, Uniformed Services University of the Health Sciences]



Diatom, a Unicellular Algae. The cell walls of diatoms are made of silica and come in a variety of shapes. These microscopic algae may be either fresh or saltwater, are photosynthetic, and play a role in carbon cycling.
[© Stanley Flegler / Visuals Unlimited]

- Microbial communities can extend in size from cubic millimeters (or smaller) to cubic kilometers. Even relatively simple communities can have millions of genes, giving them a genetic diversity substantially greater than that of higher life

forms, even humans. Recent investigations have focused on collecting DNA fragments from environmental samples in the sea and

other natural ecosystems. These “metagenomics” studies have given us a glimpse into the intricacies of these natural ecosystems and their diverse functions.

References are listed on the last page.

GTL RESEARCH STRATEGY

The GTL research program uses microbial DNA sequences as the foundation upon which to build an integrated view of how life works. This whole-systems approach envisions microbes as a complete set of intersecting processes and will illuminate fundamental principles of life and move us closer to practical applications.

Multiscale Explorations

Obtaining this level of knowledge requires investigations of both individual microbes and their natural communities. GTL analyzes key microbial properties and processes on three levels.

- **Molecular:** Focusing on genes, proteins, multicomponent protein complexes, and other biomolecules that perform cellular functions.
- **Whole cell:** Investigating how molecular processes, networks, and subsystems are controlled and coordinated.
- **Microbial community:** Understanding how microbes interact to carry out coordinated ecosystem processes.

The GTL research strategy will rely on traditional DOE strengths in advanced scientific technologies including computing and on multidisciplinary teams focused on strategic science goals and managed for results.

New User Facilities, Integrated Computational Environment

While individual microbes are among the simplest of organisms, their species diversity and community interactions require analyses of great scale and complexity. To meet these challenges, technologies will be scaled up in new high-throughput user facilities (see User Facilities, p. 5) that will support the comprehensive level of analysis required for complete systems knowledge. An integrated computational environment will link all data

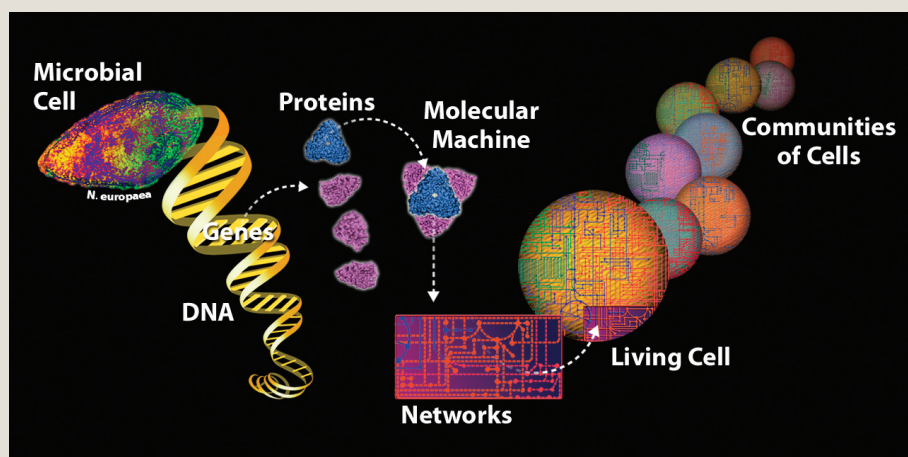
with theory, modeling, simulation, and experimentation to derive principles and develop and test biosystems theories.

GTL Knowledgebase

Computational comparison of DNA sequences across species has become a powerful analytical technique, yielding insights into gene function and facilitating hypothesis development. In the new era of systems biology, all-against-all comparisons of the much more extensive data amassed in GTL—all linked to DNA sequence—will sharpen and accelerate insights into microbial processes and research strategies. The GTL computational knowledgebase, along with high-throughput facilities, ultimately will reduce analysis time for microbial systems from many years to months, bringing with it more timely applications to national problems.

The Basics: From DNA to Living Cells to Communities

- Cells contain DNA—the hereditary material of all living systems.
- The genome is an organism's complete set of DNA.
- DNA contains genes whose sequences specify how and when to build proteins.
- Proteins perform most essential life functions, often working together as molecular machines. In addition, they form most cell structures.
- Molecular machines interact through complex, interconnected pathways and networks to make the working cell come alive.
- Communities of cells are associations of microbes (each a single cell) working together in a particular environmental niche.



N. europaea: M. A. Bruns, Center for Microbial Ecology, Michigan State University

USER FACILITIES

Technical challenges presented by GTL analyses and the scale of systems that must be understood—from genomes to ecosystems—exceed current capabilities. As we have learned from the genome projects, consolidating technologies and focusing on aggressive goals will drive dramatic improvements in performance, quality, and cost, ultimately accelerating discovery and reducing the time needed to develop useful applications.

Building on this experience, DOE's Office of Science has proposed four new facilities to make GTL goals tractable and affordable and to support the next generation of industrial biotechnology. Although each facility can serve a user community for a wide range of independent studies, the entire suite has complementary strengths that together can help provide complete systems knowledge.

Two focus on analysis of properties and functions of cellular components, proteins, and molecular machines:

- **The Facility for Production and Characterization of Proteins and Molecular Tags** will produce all proteins encoded in any genome; create molecular tags that allow

"Many researchers are stymied by lack of access to the expensive instruments that would enable them to make the greatest strides." — M. Schaechter, R. Kolter, and M. Buckley, *Microbiology in the 21st Century: Where Are We and Where Are We Going?* American Society for Microbiology (2004).

each protein to be identified, located, and manipulated in living cells; and perform biochemical and biophysical characterization of proteins produced.

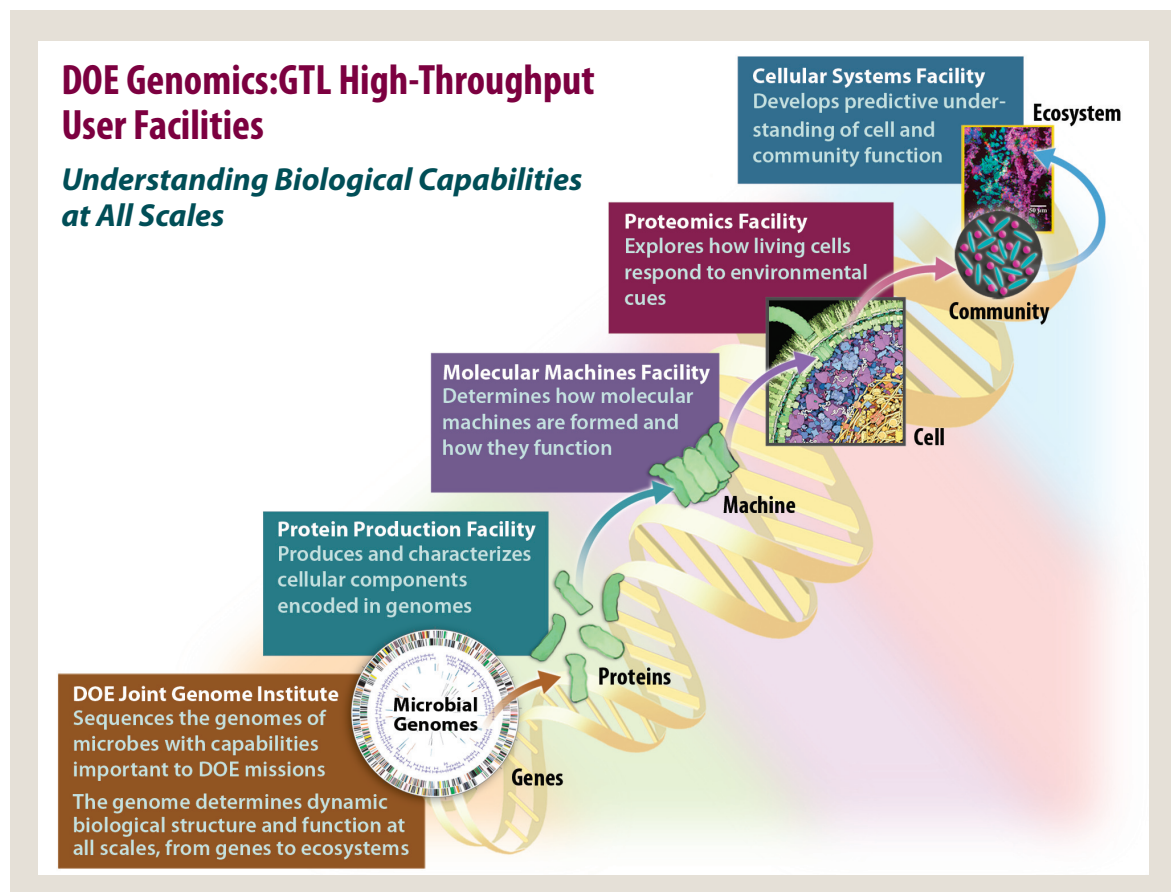
- **The Facility for Characterization and Imaging of Molecular Machines** will identify and characterize molecular assemblies and interaction networks to determine how cellular molecular processes function on a whole-systems basis.

The other two will concentrate on analyzing microbial system responses and functions at the molecular, cellular, and community levels:

- **The Facility for Whole Proteome Analysis** will identify (1) all proteins and other molecules that a microbe or community creates under controlled conditions and (2) key pathways and other processes.
- **The Facility for Modeling and Analysis of Cells and Communities** will concentrate on the systems-level

study of living cells in complex and dynamic structured communities.

These four facilities will bring together the biological, physical, computational, and engineering sciences to create a new infrastructure for biology and the industrial biotechnology needed to solve 21st Century problems. DOE's technology programs can work with industry to apply such capabilities and knowledge to a new generation of processes, products, and industries.



GTL RESEARCH IN PROGRESS

Numerous projects funded by BER since FY 2002 have established a strong foundation for GTL. Projects focus on systems biology, with some pilots chosen to establish capabilities and concepts for later scaleup in user facilities. GTL research milestones and brief descriptions of selected projects follow.

GTL Research Milestone 1: Develop Techniques to Determine the Genome Structure and Functional Potential of Microbes and Microbial Communities

- DOE has long supported the sequencing of individual microbial genomes. Recent studies are focusing on sequencing whole microbial communities in diverse natural environments in order to explore the capabilities of microbes that cannot be cultured. Such “environmental genomics” approaches have been applied to studying communities in a specific locale as well as single genes, pathways, and whole organisms. Data analyses have revealed a broad spectrum of genomes, genes, and previously undiscovered functions. These studies will produce a multitude of new insights into the dynamics between microbes and their environments and potentially could catalyze development of numerous practical applications.
- Improved methods for synthesizing genomes are being developed to test the understanding of gene function and regulation.
- Technologies are being refined, validated, and deployed in increasingly automated pilot pipelines focusing on a number of microbes to cultivate, isolate, stabilize, and characterize molecular complexes by making use of miniaturization and other

developments. Early results underscore the need for multiple technologies to characterize the thousands of protein complexes required by GTL studies each year and to eliminate the many bottlenecks that remain.

GTL Research Milestone 2: Achieve a Systems-Level Understanding of Microbial Cell and Community Function, Regulation, and Dynamics

- A number of in-depth systems biology projects on individual organisms focus on integrating the analyses of cellular proteomes, biochemistry, and imaging and on modeling pathways.
- Research on gene regulation is exploring master regulators, which tightly and predictably control cellular processes and responses to internal and external signals.
- Projects investigating biological mechanisms with potential for alternative fuel production include studies on the role of cellulose-binding modules in cellulolytic activity (for cellulosic ethanol) and large-scale analysis of genes and metabolic pathways involved in photolytic hydrogen production.

Research at the Interface of Disciplines Brings Synergistic Impacts

“As we continue to gain the ability to work at very small scales and to probe the dynamic, three-dimensional structure of molecules, the interface between physical science and life science is of critical importance. Work at the interface of frontier disciplines like bioinformatics, genomics, proteomics, and nanotechnology is greatly enhanced by DOE’s capacity in large-scale computation and research tools based upon physical science.” — *Critical Choices: Science, Energy, and Security: Final Report of the Secretary of Energy Advisory Board’s Task Force on the Future of Science Programs at the Department of Energy*, October 13, 2002.

Joint Genome Institute

The sequences of DOE-relevant microbes have been provided largely by BER’s Joint Genome Institute (JGI), an important resource that is producing microbial, microbial-community, and other genome sequences. As a user facility, JGI also makes its services available to the broader community. To help researchers analyze the deluge of DNA data on microorganisms, JGI recently brought online a clearinghouse web site, Integrated Microbial Genome (<http://img.jgi.doe.gov/v1.1/main.cgi>).

GTL Research Milestone 3: Develop the Knowledgebase, Computational Methods, and Capabilities to Advance Understanding and Prediction of Complex Biological Systems

- A computational framework is being formulated for comparative analysis of functional genomic data and computational models.
- Computational methods are being devised to predict the wiring diagrams of various microbial response networks, which comprise signaling, regulatory, and metabolic components.

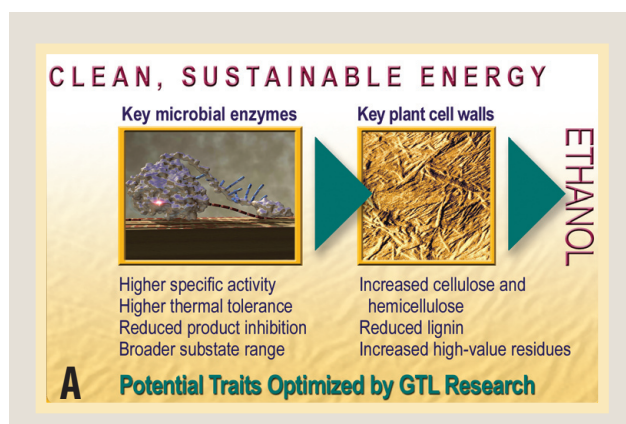
POTENTIAL GTL APPLICATIONS

GTL capabilities ultimately will contribute to dramatically shortening the technology cycle, allowing frontier science to be incorporated more directly into useful systems and reducing time and costs expended between discovery and use. Comprehensive GTL data and models will allow incorporation of scientific discoveries at molecular-level time and spatial scales into larger models and simulations. These models will cover the large process, spatial, and time scales used in mission applications for systems engineering of application technologies (e.g., biofuel production and bioremediation) and policy-support products (e.g., climate models, economic models, and integrated assessments). Some potential applications are described below.

Cellulose-Derived Ethanol

Ethanol from corn already is used as a substitute or octane booster for gasoline, displacing about 2% of fossil-fuel use. Ethanol from this source, however, has limited potential for substantial improvements in volume and cost. Biotechnology offers the promise of dramatically increasing ethanol production using cellulose from plant waste or high-biomass crops grown for energy uses.

Using cellulose—the most abundant biological material on earth—and other constituents of plant cell walls as feedstocks can result in improved energy yields, compared with cornstarch. The current commercial approach for converting cellulose into ethanol, however, uses a complicated and expensive multistep process. GTL research will lead to an understanding of the processes involved in bioethanol production, increased production efficiencies, and reduced costs and net carbon dioxide emissions from the transportation sector. In the coming decades, this R&D will help bioethanol become a cost-competitive alternative to gasoline (see Fig. A).

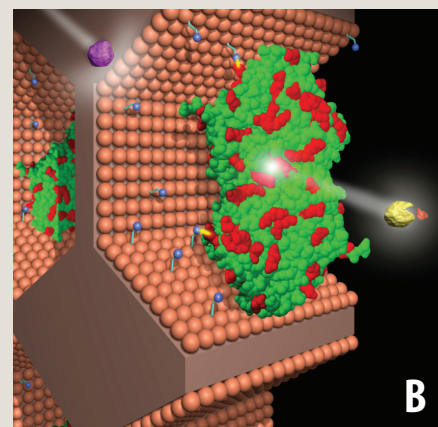


Synthetic Nanostructures

Understanding the sophisticated biochemistries of microbes can lead to the discovery of ways to isolate and use their components to carry out some of the functions of living cells.

An example in Fig. B shows the enzyme organophosphorus hydrolase (OPH, green), which has been embedded in a synthetic nanomembrane (mesoporous silica) that enhances its activity and stability [*J. Am. Chem. Soc.* **124**, 11242–43 (2002)]. The OPH transforms toxic substances (purple molecule at left of OPH) to harmless by-products (yellow and red molecules at right). Applications such as this could optimize the functionality of countless enzymes for efficient production of energy, removal or inactivation of contaminants, and sequestration of carbon to mitigate global climate change. The knowledge gained from GTL also could be highly useful in food processing, pharmaceuticals, separations, and the production of industrial chemicals.

Enhancing Enzymes



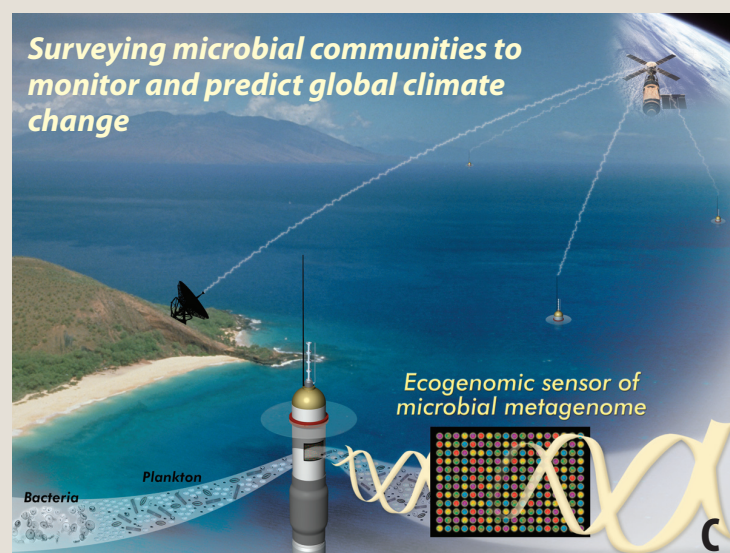
E. Ackerman, Pacific Northwest National Laboratory

Ecogenomic Sensors

Nanoscale environmental genomic sensors (see Fig. C) may one day be used to monitor microbial populations

—continued, see *Applications*, p. 8

Taking the Pulse of the Ocean



GTL PROGRAMMATIC BACKGROUND

GTL is jointly sponsored by the Office of Biological and Environmental Research (BER) and Office of Advanced Scientific Computing Research, both within the DOE Office of Science (SC). GTL also is coordinated with other relevant DOE programs to facilitate the transfer of new resources, tools, and processes to the marketplace.

Scientific and technological progress achieved during the Human Genome Project, initiated in 1986 by DOE, provided the foundation for launching the GTL program in 2002. This program brings together biologists, physical scientists, and computing scientists to generate the scientific understanding needed to develop biobased products for energy, environmental, and climate needs. The 2005 *GTL Roadmap*, which builds on and expands the initial program plan, is the result of 3 years of collaboration among hundreds of scientists and technologists. Many genome sequences used in GTL and determined by BER programs have made important contributions to the understanding of biology, genetics, and evolution.

For more than half a century, SC has envisioned, designed, constructed, and operated many of the world's premier physical research facilities, which continue to grow in importance to biology. Through its strategic approach to research and through the proposed GTL facilities, DOE's strengths in the biological, physical, and computational sciences will bring breakthrough technologies to bear on biology. This multidisciplinary approach is consistent with the DOE research tradition. GTL user facilities are among those featured in SC's 20-year plan, with the SC director noting that "Investment in these [GTL] facilities will yield extraordinary scientific breakthroughs and vital societal and economic benefits" (*Facilities for the Future of Science: A Twenty-Year Outlook*, www.sc.doe.gov).

References, continued from p. 2-3

1. T. P. Curtis and W. T. Sloan, "Prokaryotic Diversity and Its Limits: Microbial Community Structure in Nature and Implications for Microbial Ecology," *Curr. Opin. Microbiol.* 7, 221-26 (2004).
2. V. Torsvik and L. Øvreås, "Microbial Diversity and Function in Soil: From Genes to Ecosystems," *Curr. Opin. Microbiol.* 5(3), 240-45 (2002).
3. *Bergey's Manual*, 2nd edition (2001).
4. T. S. Bibby et al., "Low-Light-Adapted *Prochlorococcus* Species Possess Specific Antennae for each Photosystem," *Nature* 424, 1051-4 (2003).
5. E. V. Armbrust et al., "The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism," *Science* 306, 79-86 (2004).
6. J. C. Venter et al., "Environmental Genome Shotgun Sequencing of the Sargasso Sea," *Science* 304, 66-74 (2004).
7. The News Staff, "Breakthrough of the Year: The Runners Up," *Science* 306, 2013-17 (2004).
8. T. C. Onstott et al., "Indigenous and Contaminant Microbes in Ultradeep Mines," *Environ. Microb.* 5(11), 1168-91 (2003); K. Kashefi and D. R. Lovley, "Extending the Upper Temperature for Life," *Science* 301, 934 (2003).

Applications, continued from p. 7

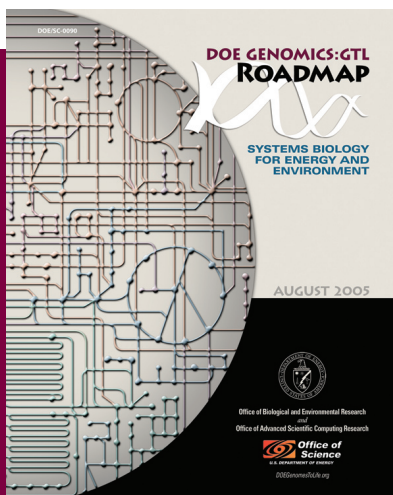
and their interactions with environmental processes, including those affected by climate change. The real-time approach envisioned by DOE for the National Oceanographic Partnership Program merges information from genome research programs with nanotechnologies and smart sensors.

The knowledge gained will enhance understanding of the genetic diversity and functions of microbial communities and help answer key questions about their influence on ocean and terrestrial biogeochemical cycles. Microbial sentinels of ecosystem changes may forewarn the approach of such events as red tide caused by an increase in *Pfiesteria* species.

Web Sites for More Information and to Request Copies of the Roadmap

DOE Office of Science genome programs:
doegenomes.org

GTL roadmap:
doegenomestolife.org/roadmap/



Contacts

David Thomassen (301.903.9817)
U.S. Department of Energy (SC-23)
Office of Biological and Environmental
Research
david.thomassen@science.doe.gov

Gary Johnson (301.903.5800)
U.S. Department of Energy (SC-21.1)
Office of Advanced Scientific Computing
Research
gary.johnson@science.doe.gov